

Tomato Phenylalanine Ammonia-Lyase Gene Family, Highly Redundant but Strongly Underutilized*

Received for publication, June 9, 2008, and in revised form, October 2, 2008. Published, JBC Papers in Press, October 6, 2008, DOI 10.1074/jbc.M804428200

Ancheol Chang, Myung-Ho Lim, Shin-Woo Lee¹, E. Jane Robb, and Ross N. Nazar²

From the Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Phenylalanine ammonia-lyase (PAL) is an important enzyme in both plant development and pathogen defense. In all plants it is encoded by a multi-gene family, ranging in copy number from four in *Arabidopsis* to a dozen or more copies in some higher plants. Many studies indicate that alternate genes are differentially regulated in response to environmental stimuli. In this study, Southern blot and dot blot analyses in tomato indicate a surprisingly large family of related sequences with ~26 copies in the diploid genome, some easily distinguished by restriction enzyme digestion. Analyses of a BAC genome library suggest that the genes are generally not clustered. A more detailed comparison of the gene sequences using PCR to isolate the individual copies and reverse transcription-PCR to study the transcripts that they encode indicates a significant diversity in the gene sequences themselves, but surprisingly only one mRNA transcript can be detected even when additional expression is induced by pathogen growth or wounding. Consistent with previous reports in other plants, a parallel study with a closely related plant, the potato, indicates a much broader utilization of the PAL genes, highlighting the unusual nature of this family in tomato and of the mechanism(s) that silences so many members. Plant transformation analyses further demonstrate the presence of very active silencing, suggesting aggressive competition between PAL gene duplication and copy inactivation during PAL gene evolution.

The duplication of individual genes, chromosome segments, and even entire genomes has long been considered an important source of evolutionary change leading to new gene functions (1–4). For example, in the *Arabidopsis* or rice genomes, up to 90% or 62% of loci are duplicated, respectively (3). Despite this, it is much less clear how duplicated genes evolve to serve new functions. Presumably at least one copy is stable and maintains the existing function(s), whereas an expendable copy undergoes changes that are preserved by natural selection (neofunctionalization). Partial compromise of function (subfunctionalization) through mutation accumulation as well as stochastic silencing of such genes also are thought to play sig-

nificant roles in the passive origin of new species (2). How such events might occur, however, remains an area of considerable debate (2).

Phenylalanine ammonia-lyase catalyzes the conversion of L-phenylalanine into *trans*-cinnamate, the initial committed step of the multi-branched phenylpropanoid pathway in higher plants. As the first step in phenolic metabolism, this is a key biochemical reaction in both plant development and defense. In all studied plants PAL³ protein is encoded by a multi-gene family ranging from a few members in many species such as raspberry or bean (5, 6) to a dozen or more copies in others such as potato (7). In tomato, hybridization analyses based on fragment length polymorphism initially indicated at least five different classes of PAL genes (8). One of these classes (PAL 5) was distinctly (5–6-fold) more common. This type of PAL gene sequence also was found to be strongly expressed (9). Studies have shown that in at least one member of this group, transcription is initiated from two sites that appear to be differentially regulated (9) in response to changes in light or wounding or to infection by a plant pathogen.

The high copy number associated with the PAL gene family in tomato together with a readily demonstrated and wide sequence polymorphism appears to make this gene family an attractive model for studies of gene duplication, evolution, and silencing. The pattern of sequence polymorphism suggests a highly duplicated family with both functional and more divergent members that might be responsive to various developmental events and environmental stresses. Indeed, initial studies on several of the divergent members indicate that dramatic changes are imposed by premature termination codons and frameshift mutations (8). At least one of these (PAL 1) appeared to be slightly expressed, which raises intriguing questions about altered function (8).

To more precisely define the number and nature of the tomato PAL gene family members and to further assess the degree of their expression, studies have been undertaken to determine more accurately the actual copy number in tomato and the significance of sequence heterogeneity with respect to gene expression. Gene cloning has been used to determine the genomic sequences and detect their actual rRNA transcripts. The comparisons indicate a very redundant gene family comprising at least 18 different sequences, but almost all appear silenced. Only a single sequence represents all or an overwhelming majority of the transcripts.

* This work was supported by the Natural Sciences and Engineering Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ Present address: Dept. of Crop Science and Biotechnology, JinJu National University, JinJu 660-758, South Korea.

² To whom correspondence should be addressed: Dept. of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Fax: 519-837-2075; E-mail: rnnazar@uoguelph.ca.

³ The abbreviations used are: PAL, phenylalanine ammonia-lyase; RT, reverse transcription.

EXPERIMENTAL PROCEDURES

Plant Material—Tomato (*Lycopersicon esculentum* L.) cv. Craigella near-isolines, susceptible (CSG CR 26) and resistant (CRG CR 218) to *Verticillium dahliae* race 1, were grown in growth chambers (Convion) at 26 °C for 14 h in the light and at 22 °C for 10 h in the dark. Before planting in a mixture of Pro-mix, vermiculite, and surface media (3:2:1), the seeds were surface-sterilized in 1% hypochlorite solution for 10 min and then rinsed in distilled water three times for 1 h. The plants were transferred after seedling germination to 17 × 13 × 6-cm Kord cell flats (6 plants/flat). The soil was kept moist at all times. Hoagland's solution was used as fertilizer once a week (10).

Preparation of Tomato Nuclear DNA—The nuclei were isolated from young leaves of 4-week-old tomato plants using a method similar to that described by Cushman (11). 8 g of fresh leaves or other plant tissue were divided into four batches. Each batch was finely chopped and homogenized in 10 ml of nuclei isolation buffer I containing 250 mM sucrose, 3 mM CaCl₂, and 6 mM Tris-HCl, pH 7.2) using an Omni mixer (Omni International, Inc. Marietta, GA) with five bursts of 30 s at speed 3. The homogenate was then filtered through three layers of cheese-cloth, and 10% Triton X-100 solution was added dropwise to the filtrate to give a final concentration of 0.3% to lyse the chloroplasts. After 5 min of incubation on ice, nuclei isolation buffer II containing 1.25 M sucrose, 3 mM CaCl₂, and 6 mM Tris-HCl, pH 7.2, was added to obtain a final concentration of 0.75 M with respect to sucrose and mixed by gentle swirling. The nuclei were collected by centrifugation at 700 rpm (57 × g) for 2 min at 4 °C in a Beckman JS7.5 swinging bucket rotor. The decanted supernatant was respun at 1800 rpm (380 × g) for 10 min at 4 °C in the same swinging bucket rotor. Each pellet was gently suspended in 250 μl of nuclei isolation buffer I and collected in a tube. An equal volume of nuclear lysis buffer (2 M NaCl, 10 mM EDTA, 0.6% SDS, and 0.1 M Tris-HCl, pH 7.5) preheated to 60 °C was added, and the suspension was incubated for 10 min at 65 °C with occasional shaking. The nuclear lysate was then subjected to chloroform extraction, and the aqueous phase was mixed by gentle swirling with an equal volume of 95% ethanol containing 2% potassium acetate. The DNA was spooled out with a clean glass rod and dissolved in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Finally, the nuclear genomic DNA was purified further using CsCl density gradient centrifugation (12).

Preparation of Cellular RNA—Total cellular RNA was prepared from tomato tissues using a method similar to that described by Heinz *et al.* (13). Tomato leaves or other tissues (0.5 g) were collected by cutting with scissors, immediately frozen, and ground in liquid nitrogen using a mortar and pestle. The powder was transferred to a homogenization tube containing 5 ml of extraction buffer (0.14 M NaCl, 0.05 M sodium citrate, 0.3% SDS, pH 6.8) and 5 ml of phenol and was homogenized for 10 strokes. The homogenate was transferred to a 10-ml plastic tube and incubated at 65 °C for 10 min with occasional shaking. After being cleared by centrifugation at 10,000 rpm for 10 min at 15 °C in a Beckman JA20 rotor, the aqueous phase was collected and precipitated with 2 volumes of ethanol containing 2% potassium acetate for 3 h or longer at -20 °C.

After centrifugation at 10,000 rpm for 20 min at 4 °C in the same rotor, the pellet was washed with ethanol, dried, resuspended in 300 μl of sterilized distilled water, transferred to a microcentrifuge tube, and reprecipitated with a one-half volume of 7.5 M ammonium acetate and 2 volumes of 95% ethanol. The RNA was further collected by centrifugation for 3 min in a microcentrifuge and washed with ethanol. Finally, the pellet was dissolved in 200 μl of sterilized distilled water. The amount and quality of RNA was determined by the absorbance at 260/280 nm and confirmed by gel analysis of the ribosomal RNA components after being stained with methylene blue (12).

PCR Amplification and Determination of PAL Gene Sequences—Targeted genome sequences were amplified using either *Taq* or *Pfu* DNA polymerase. For *Taq* DNA polymerase (MBI Fermentas, Hanover, PA), 1 unit of enzyme was added to 50 μl of reaction mixture containing 50 mM KCl, 1.5 mM MgCl₂, 10 mM Tris-HCl (pH 9.0 at 25 °C), 0.1% Triton X-100, 100 μg/ml bovine serum albumin, 0.2 mM of each deoxyribonucleotide triphosphate, 15–20 pmol of each gene-specific primer, and 200–300 ng of nuclear genomic DNA. For the high fidelity polymerase, 1.25 units of *Pfu* DNA polymerase (MBI Fermentas) was added to 50 μl of reaction mixture containing 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 20 mM Tris-HCl, pH 8.2, 0.1% Triton X-100, 0.1 mg of bovine serum albumin, 0.1–0.2 μM of each gene-specific primer, and 200–300 ng of nuclear genomic DNA. Two coding regions of the PAL genes, major portions of exon I and exon II, respectively, were targeted using alternate primer sets, a forward primer, 5'-ACAAATGGACA-TRTTAAT-3' (where R is A or G; P1) and 5'-CTTCTATGAGATGTTGC-3' (reverse primer; P2) for the first exon, and 5'-GAATTCAGTACTATT-3' (forward primer; P3) and 5'-GATAGGTTGATGACATT-3' (reverse primer; P4) for the second exon, respectively. Targeted fragments were amplified in a programmable heating block (Gene ATAQ Controller; Pharmacia LKB) using 30 reaction cycles consisting of a 30-s denaturation step at 95 °C, a 1-min annealing step at 43 °C, and a 1-min elongation step at 72 °C.

After amplification, PCR products were purified by chloroform extraction and initially assessed by agarose gel electrophoresis (12). Each amplified fragment was subcloned into the pTZ19R cloning vector, first digested with restriction endonuclease *Sma*I and then "tailed" with thymidylic acid residues at the 3' ends of the *Sma*I site as previously described (14). When *Pfu* DNA polymerase was used, the amplified products were first "tailed" with adenylic acid residues. After incubation with T4 DNA ligase overnight at 14 °C, the recombinant DNA was used to transform *Escherichia coli* DH5α and ampicillin-resistant colonies were then screened by colony hybridization (12). Each colony grown on an LB agar plate containing 100 μg/ml of ampicillin was transferred to a nitrocellulose filter (Whatman, Brentford, UK) using a sterilize toothpick and then grown on a fresh LB plate containing 100 μg/ml of ampicillin for 6 h at 37 °C. Each filter containing the bacterial colonies was soaked in 0.5 M NaOH for 8 min twice and neutralized with 0.5 M Tris-HCl (pH 7.5) and 1.5 M NaCl for 8 min. The air-dried filter was washed twice with 95% ethanol, dried at 80 °C for 2 h to fix the bacterial DNA to the filter, and then hybridized with probe solution (50% formamide, 5× SSC, 0.5% SDS) containing

labeled PAL gene-specific probes at 50 °C for 16–20 h. After hybridization, the filter was washed three times, 15 min each, with washing solution (probe solution without probes) at 50 °C and twice, 5 min each with 2× SSC at room temperature. Finally, the air-dried filter was exposed to x-ray film (Kodak XAR-5) at –70 °C for 12–16 h using an intensifying screen. To prepare probes, PCR-amplified DNA fragments for the two PAL protein-encoding regions were purified from agarose gels using a gel extraction system (Qiaquick gel extraction kit; Qiagen), denatured, and labeled with the large fragment of DNA polymerase I (Klenow; MBI Fermentas), using random hexamers as primers and [α -³²P]dATP as previously described (15).

For DNA sequence analyses, plasmid DNAs were prepared using a miniprep DNA purification system (Wizard Plus SV Minipreps, Promega). Both strands of the insert in each clone were sequenced with either a forward primer, 5'-GACGTTG-TAAAACGACG-3' (U2) or a reverse primer, 5'-CAGGAAA-CAGCTATGAC-3' (ARP), in an automated sequencing system (CEQ8000 8-capillary Genetic Analysis System; Beckman). All of the sequence alignment was carried out using the BioEdit Sequence Alignment Editor, version 4.3.2 (North Carolina State University).

RT-PCR Amplification and Determination of PAL mRNA Sequences—DNA for mRNA sequence analyses was prepared from cellular RNA extracts by RT-PCR amplification as described by Gerad (16). The cDNA was prepared using 20- μ l reaction volumes containing 300–500 μ g of total RNA, 2–5 pmol of a PAL gene-specific primer, 0.2 mM of each deoxyribonucleotide triphosphate, reaction buffer (5× buffer containing 375 mM KCl, 15 mM MgCl₂, 250 mM Tris-HCl, pH 8.3), 10 mM dithiothreitol, and 200 units of Moloney murine leukemia virus reverse transcriptase (Superscript II; Invitrogen). A 13- μ l aliquot containing a gene-specific primer; 5'-CTTCTATGAGATGTTGC-3' for the exon 1 region or 5'-GATAGGTTGATGACATT-3' for exon 2, RNA extract, and dNTPs was heated at 65 °C for 5 min and quickly chilled on ice. After brief centrifugation, reaction buffer and dithiothreitol were added, and the mixture was further incubated at 42 °C for 2 min. Finally, reverse transcriptase was added, and the solution was mixed by pipette before incubation at 42 °C for 50 min. The reaction was heat-inactivated at 75 °C for 5 min, and 2 μ l of the product were used as a template for subsequent PCR amplifications. The cDNAs were PCR-amplified using the exon 1- and 2-specific primer pairs; the products were cloned subsequently, and their sequences were determined as described for the PAL genes.

Agrobacterium-mediated Tomato Transformation—Tomato cells were transformed with *Agrobacterium* containing pCAMP5GUS recombinants as described by Fillatti *et al.* (17), based on an online protocol with minor modifications. Briefly, tomato (cv. Craigella, susceptible) seeds were sterilized with 70% ethanol for 1 min followed with 5.25% NaOCl for 20 min with swirling and then rinsed six times with sterile distilled water. The sterilized seeds were transferred to culture vessels containing 50 ml of medium A for cultivation under light (25 °C, 16 h photoperiod). Eight- or 9-day-old seedlings were placed in sterile distilled water, and the cotyledons were cut with sterile scissors and transferred to agar plates containing 20

ml of medium B with 100 μ M of acetosyringone (18). The plates were sealed with parafilm and precultured for 24 h at 25 °C in the dark. Cultures of *Agrobacterium* containing a transforming plasmid construct (pCAMP5GUS) were grown overnight in 6 ml of YEP medium with 25 μ g/ml streptomycin and 50 μ g/ml kanamycin with shaking at 28 °C. The overnight culture was diluted with liquid Murashiga Skoog medium to an absorbance at 600 nm of 0.5, and 8 ml were poured onto the cotyledons placed in Petri dishes with liquid medium B containing 100 μ M of acetosyringone. After incubation for 30 min with occasional swirling, the excess *Agrobacterium* suspension was removed by aspiration, and the cotyledons were transferred to preculture plates containing medium B with 100 μ M of acetosyringone for co-cultivation at 25 °C for 48 h in the dark.

The co-cultured cotyledon explants were transferred to Petri dishes with medium C containing 100 μ g/ml kanamycin and 150 μ g/ml timentin and subcultured on fresh medium C every 3–5 days to initiate calli. After 2–3 weeks, calli with shoot primordia were excised as small pieces and transferred to regeneration and selection medium D for weekly subculturing. When shoot stems had elongated to 2–4 cm, the shoots were excised from the calli and transferred to rooting medium E in magenta boxes. Plantlets of about 5-cm height subsequently were removed from the culture vessels and transplanted into regular pots with moistened soil, a mixture of Pro-mix, vermiculite, and turf (3:2:1). The pots were covered with transparent plastic wrap before being moved to a controlled growth room (25 °C, 16-h photoperiod). After 3 days, the holes were made in the plastic wrap cover that were enlarged further everyday to adapt the plants gradually and avoid desiccation. Transgenic plants were transplanted into larger pots after 2–3 weeks.

Assay of Methylated DNA Sites—Methylated DNA sites were assayed using bisulfite conversion genomic sequencing as previously described (19). Tomato nuclear genomic DNA digested with EcoRI restriction enzyme was diluted to 2 μ g/50 μ l with double distilled water and 5.5 μ l of 3 M NaOH were added. After incubation at 37 °C for 10 min, 30 μ l of freshly prepared 10 mM hydroquinone and 520 μ l of freshly prepared 5 M sodium bisulfite were added and mixed. Mineral oil was layered over the solution before incubation at 50 °C in the dark for 9 h. After mixing with 600 μ l of miniprep neutralization solution, Wizard plus SV miniprep DNA purification system (Promega, Madison, WI), the mixture was added to a miniprep column and cleaned by microcentrifuge for 1 min. DNA in a column was then washed with 80% isopropanol and eluted with 50 μ l of double distilled water preheated to 65 °C. The eluted DNA was mixed with 5.5 μ l of 3 M NaOH and incubated at room temperature for 5 min. A one-half volume of 7.5 M ammonium acetate and two volumes of 95% ethanol were added, and the DNA was precipitated at –20 °C, overnight. Finally, the pelleted DNA was washed with ethanol and resuspended in 20 μ l of double distilled water. A 200–300-ng aliquot of bisulfite converted or unconverted control DNA was used as a template for PCR amplification of PAL gene upstream regions using primer sets designed for bisulfite converted or unconverted DNA, as appropriate. All PCR-amplified products were cloned using pTZ19R for subsequent DNA sequence analyses.

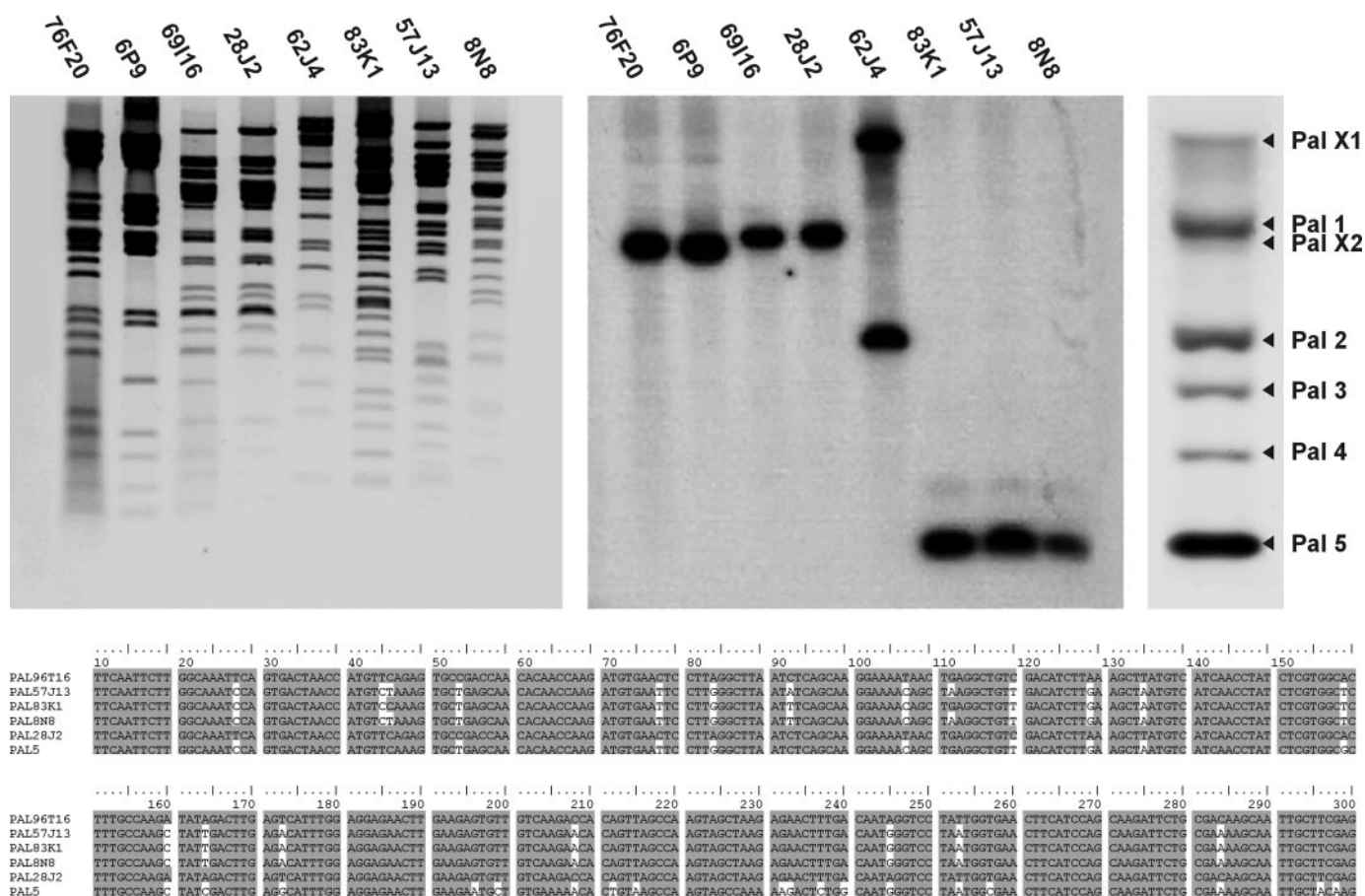


FIGURE 1. Distribution of PAL gene sequences in the tomato genome. A deep coverage tomato BAC library on nylon membranes (21) was screened with a labeled 580-bp EcoRI PAL5 gene fragment, and DNA prepared from the selected clones (76F20–8N8) was digested with EcoRI endonuclease, fractionated on a 1% agarose gel, and hybridized with the same probe after being blotted onto nylon membrane (*center panel*). Purified genomic DNA (8) also was digested, fractionated, and hybridized with the same labeled PAL5 gene fragment (*right panel*). The sequences of PAL genes identified in the library were determined, and a representative portion is compared below together with the previously reported PAL5 gene sequence; identical nucleotides are indicated by shading.

RESULTS

Previous studies on the PAL gene sequences in tomato indicated a heterogeneous family of genes, but the degree of heterogeneity, both with respect to gene sequence and expression, remained unclear. Southern hybridization analyses after EcoRI digestion have demonstrated at least five different members but did not resolve whether differences in band intensities reflected further gene heterogeneity or simply resulted from transfer/hybridization artifacts (8). More exhaustive transfers in this study actually increased the number of different bands to as many as seven, but once more, the significance of the relative intensities remained unresolved. As shown in Fig. 1 (*right panel*), when membranes were hybridized with a 581-bp EcoRI fragment of a tomato PAL5 gene (20), of the seven distinct bands most appear approximately equal in intensity; however, the shortest remained about 5-fold darker and the additional fragments (X1 and X2) appeared somewhat underrepresented.

To quantify the copy number more accurately, in the present study dot blots initially were used to determine the actual total number of PAL gene-related sequences (Fig. 2). Known amounts of a PCR-amplified PAL5 gene fragment (*upper row*), purified tomato genomic DNA (*middle row*), and cloned PALX1 plasmid DNA (*lower row*) were spotted, and the membranes again were hybridized with labeled PAL5 gene probe. As

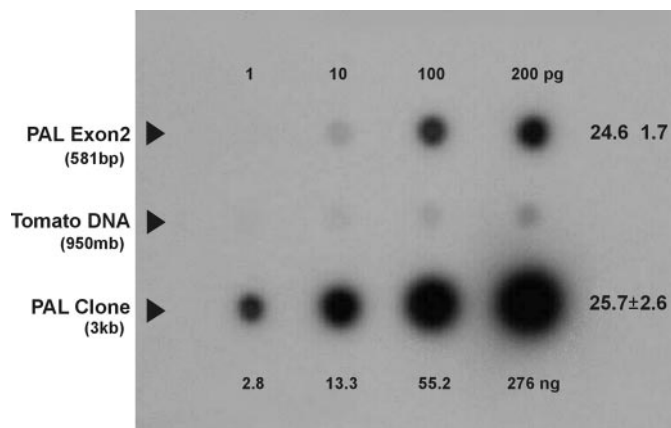


FIGURE 2. Estimated number of PAL gene sequences in the tomato genome. Increasing amounts of PCR-amplified PAL5 gene fragment (*upper row*) or a PALX1 plasmid clone (*lower row*) were blotted on polyvinylidene difluoride membrane together with tomato genomic DNA prepared from purified nuclei (*center row*) and hybridized with labeled a 580-bp EcoRI PAL5 gene fragment. The gene copy number was determined from the hybridization signal intensities using each standard independently and assuming a 950-Mb tomato genome size; the values per diploid genome are indicated on the right \pm S.D.

shown in Fig. 2, multiple applications with differing amounts of DNA were spotted to evaluate the linearity of this scale. Based on the hybridization signal intensities and assuming a 950-Mb

genome size for tomato, the blots indicated that there were ~ 13 copies of the PAL gene sequence/haploid genome or a total of 26 copies/diploid tomato cell. As indicated, both standards resulted in very similar values, 24.6 ± 1.7 and 25.6 ± 2.6 , respectively. When combined with the results shown in Fig. 1 (*right panel*), the weakest bands would be consistent with one copy per haploid genome, the darker bands with two (PAL1 and PAL2) and five (PAL5) copies, respectively.

In addition to hybridization analyses two other analytical approaches subsequently were applied to explore further the degree of heterogeneity in the tomato PAL gene family. In the first instance, a tomato genome library was examined for clusters of PAL genes. The deep coverage tomato BAC library (LE_HBa), prepared by Budiman *et al.* (21), was obtained on nylon membranes from the Clemson University Genome Institute and screened with the most abundant 580-bp EcoRI PAL5 gene fragment also used in the dot blot experiments. The selected BAC clones of the large DNA fragments making up the library subsequently were obtained from the same source and subjected to both hybridization and DNA sequence analyses. As shown in Fig. 1, to evaluate the distribution of the PAL gene sequences in these large fragments, the eight BAC clones, which were selected as positive in this manner, initially were studied further by Southern blot hybridization. Purified plasmid DNA was digested with restriction enzymes (*left panel*) and probed by hybridization (*middle panel*) using the same PAL gene probe that was utilized in the prior studies and selection of the BAC clones themselves. Based on the observed restriction fragment length polymorphism (Fig. 1, *center panel*), in all but one instance only a single family member was identified in each BAC clone. In the single exception (62J4) only, two distinct sequences (PALX1 and PAL2) were present. Subsequent sequence analyses of these clones fully supported this conclusion. As also indicated in Fig. 1 (*lower panel*), a sequence comparison clearly shows all the sequences were consistent with unique PAL genes. Taken together these analyses show that despite the high copy number, the members of this family are not highly linked or clustered but dispersed widely in the tomato genome.

In view of the family size and the varied roles that plant phenylalanine lyases can play in plant development and pathogen defense mechanisms, further studies were undertaken to link specific members of the family with individual functions in plant cells. Initially, a survey of the expressed genes was undertaken based on sequence analyses. Using primer sets complementary to two known highly conserved sequence regions, a major portion of exon 1 (360 bp) and a comparable region in exon 2 (652 bp), DNA was prepared by PCR amplification using purified genomic DNA or cDNA, itself prepared using reverse transcriptase and purified whole cell RNA (Fig. 3). In each case, the PCR-amplified DNA was cloned using the pTZ19R vector. In the first experiments, *Taq* polymerase was used to prepare the DNA, but ambiguous single base changes were commonly observed, and the experiments were repeated using the more faithful *Pfu* RNA polymerase (22). Only the second group of experiments are presented in this report, but the same general conclusion could be drawn with either polymerase. As indicated in Figs. 4 and 5 and strongly supportive of the early studies

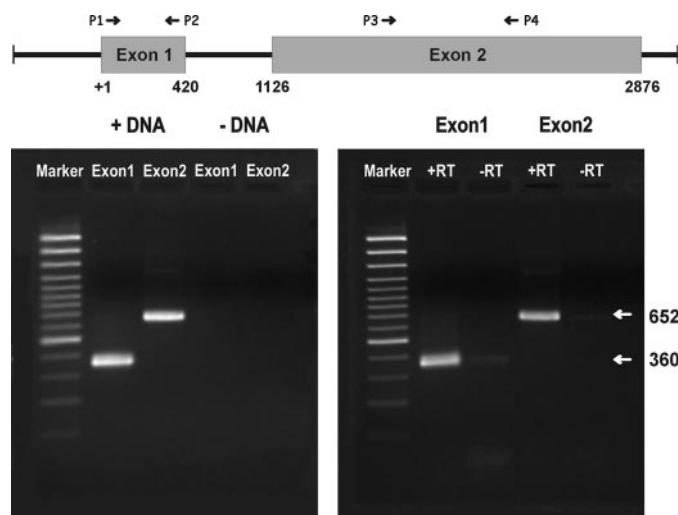


FIGURE 3. Preparation of PAL genomic or mRNA gene sequences. The represented portions of the exon 1 and 2 regions were prepared from purified genomic DNA by PCR amplification (*left panel*) or from mRNA by RT-PCR (*right panels*) using highly conserve primer sequences (P1–4) as described under “Experimental Procedures” and fractionated on 2% agarose gels. Control reactions without DNA template (–DNA) were included for the genomic DNA reactions; control reactions without reverse transcriptase (–RT) were included for mRNA. The sizes of the amplified fragments, as determined from markers, are indicated on the *right*.

using dot or southern gel blot analyses, genomic sequence analyses after DNA cloning again indicated a surprisingly large, heterogeneous family of gene members. Of the 28 clone sequences that were determined, 13 or 18 unique sequence groups were identified for the exon 1 and 2 regions, respectively. Multiple changes in sequence distinguished each group (*unshaded regions*); multiple clones sharing a common sequence were observed for four groups in exon 1 and five groups in exon 2 as indicated by the *closed boxes*.

When the same analytical approach was applied to mRNA from tomato leaves, using reverse transcriptase to first produce cDNA, the results were strikingly different. Of the 35 clones that were examined in this way (Fig. 6), again representing both exon regions, only one unique nucleotide sequence was evident. This corresponded exactly with the last group of DNA sequences for both exons. The results indicated that, despite the very large number of different PAL-encoding genes in tomato, only one of these sequences was overwhelmingly expressed. The analyses did not exclude completely the possibility of very minor amounts of other mRNA sequences, but clearly there was only one very dominant sequence being expressed in the leaves of growing tomato plants.

In view of these very unanticipated observations, two additional studies were undertaken to further evaluate the unusual nature of these results. As noted earlier, in studies of PAL-encoding genes in other plants, multiple gene families commonly have been reported (4). Expression studies in these cases often have linked different members with different tissues or environmental changes (23–28). In view of such reports, two additional comparisons were undertaken in respect to the tomato gene family. Alternate tissues or conditions were examined for alternate transcripts, and a comparable experimental strategy was applied to study the PAL gene family in a closely related plant, the potato. Both analyses again underlined a very

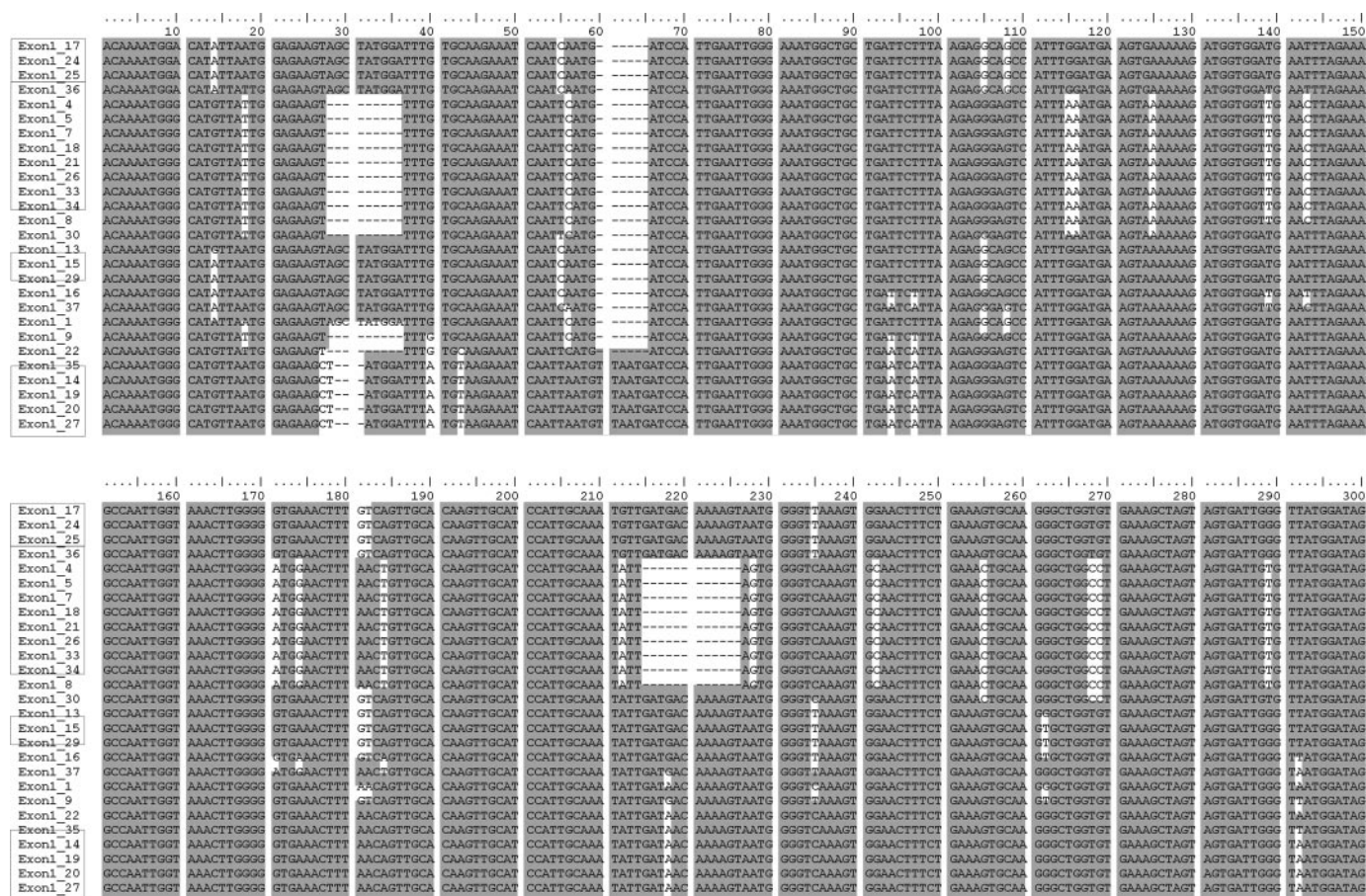


FIGURE 4. Comparison of tomato PAL gene exon 1 region sequences. Exon 1 DNA was amplified from genomic DNA as described in Fig. 3 and cloned in pTZ19R, and the sequences of randomly picked clones were determined by automated sequence analysis. Representative portions of the sequences are compared and grouped by maximum identity; groups of identical sequences are indicated by closed boxes on the left. Shared nucleotides are indicated by shading.

unusual PAL gene utilization profile in tomato. Although not as exhaustive, a study of different organs or environmental stress conditions (7) based on a comparable approach again revealed a dramatic underutilization of the PAL genes with the same dominant transcript. To facilitate a search for alternative transcripts, clones were first examined for restriction fragment length polymorphism predicted by the genomic DNA sequences (Figs. 4 and 5) using EcoRI or HinfII. As clearly evident in the examples shown in Fig. 7, this study again demonstrated that in all tomato organs and even in plants subjected to strong environmental stress associated with pathogen colonization, tissue wounding, or UV damage, there remains an equally striking underutilization of the PAL-encoding genes. Only the two HinfII digestion fragments (448 and 184 bp) predicted by the leaf-expressed sequence were evident in analytical gels (Fig. 7, right panel), and all sequences of example clones derived from callus, fruit, leaf, root, seed or stem reflected the same major transcript (Fig. 7, lower comparisons).

In view of past reports of heterogeneous PAL gene expression in other plants, a comparable study also was undertaken in potato. The high sequence conservation in this closely related plant permitted comparisons of equivalent regions. Numerous clones of both PCR-amplified genomic loci and RT-PCR-amplified mRNA again were prepared and sequenced. As summa-

rized with the example mRNA sequences tabulated in Fig. 8, the results of the studies proved very different, with significant sequence heterogeneity immediately being evident at both the DNA and RNA levels. Unlike tomato, in potato, as in other reported plant examples (7), multiple members of the PAL gene family clearly are expressed with gene utilization not being restricted to a single sequence or family member. In potato one transcript dominated, because five clones for each exon (Fig. 8, boxed) had the same sequence, but three additional sequences were readily identifiable (Fig. 8, unshaded nucleotides).

Because all of the results in tomato suggested extensive gene silencing, an attempt was made to evaluate the plant response to a new and active PAL gene promoter. As indicated in Fig. 9, a Ti plasmid-based strategy was adopted, and a chimeric gene construct was prepared containing a tomato PAL5 gene promoter (8) fused to a GUS reporter sequence and the polyadenylation site of a nopaline synthetase gene (29) in the pCAM-BIA2300 plant transformation binary vector. The construct was used to transform streptomycin-resistant *Agrobacterium*, LBA4404, by freeze-thaw (30) and cotransformed into the resistant pAL4404 virulence plasmid (31). A transformed *Agrobacterium* suspension was used to transform tomato cotyledons as described by Fillatti *et al.* (17). Regenerated transformed plants were detected by PCR amplification using nuclear DNA isolated from the leaves and primers, specific for the PAL gene

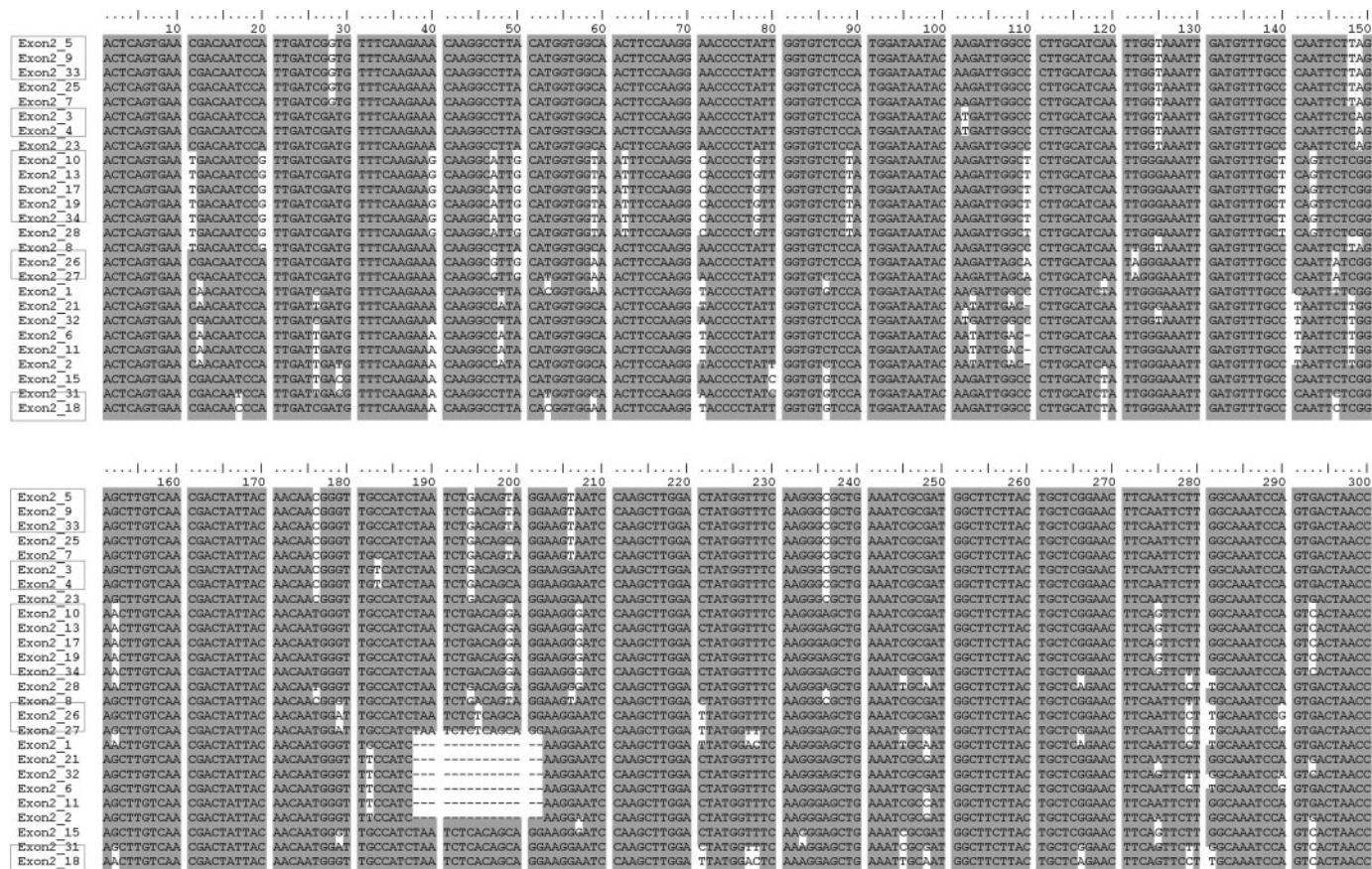


FIGURE 5. **Comparison of tomato PAL gene exon 2 region sequences.** Exon 2 DNA was amplified from genomic DNA as described in Fig. 3 and cloned in pTZ19R, and the sequences of randomly picked clones were determined by automated sequence analysis. Represented portions of the sequences are compared and grouped by maximum identity; groups of identical sequences are indicated by closed boxes on the left. Shared nucleotides are indicated by shading.

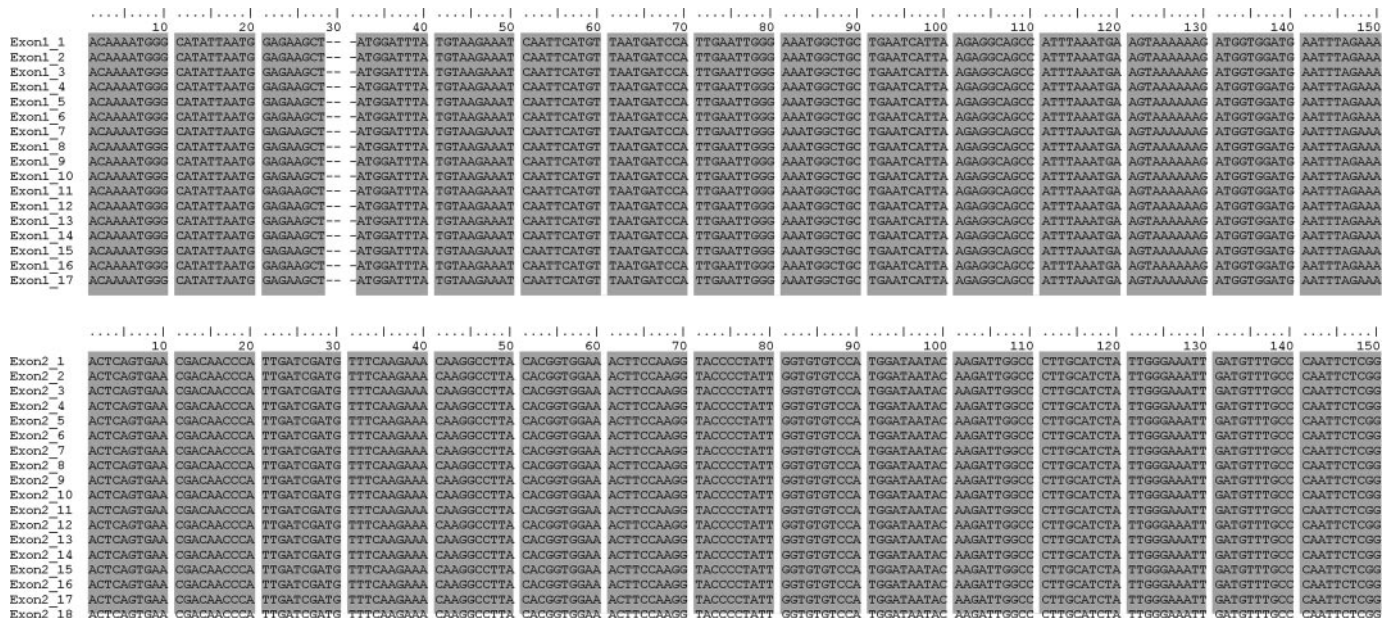


FIGURE 6. **Comparison of tomato leaf PAL mRNA sequences.** Exon 1 or 2 cDNA was prepared from tomato leaves mRNA as described in Fig. 3 and cloned in pTZ19R, and the sequences of randomly picked clones were determined by automated sequence analysis. Representative portions of the sequences are aligned to indicate the complete identity.

(5'-CACTAATCATAGTTCACAAC-3') and the GUS gene (5'-CTCCATCACTTCCTGAT-3') sequences. Endogenous tomato PAL genes also were amplified from nontransgenic

well as transgenic plants as positive controls; the two fragments (460 bp for the transgene and 652 bp for the host gene) were distinguished readily by gel electrophoresis (Fig. 9, left gel).

Tomato PAL Gene Family

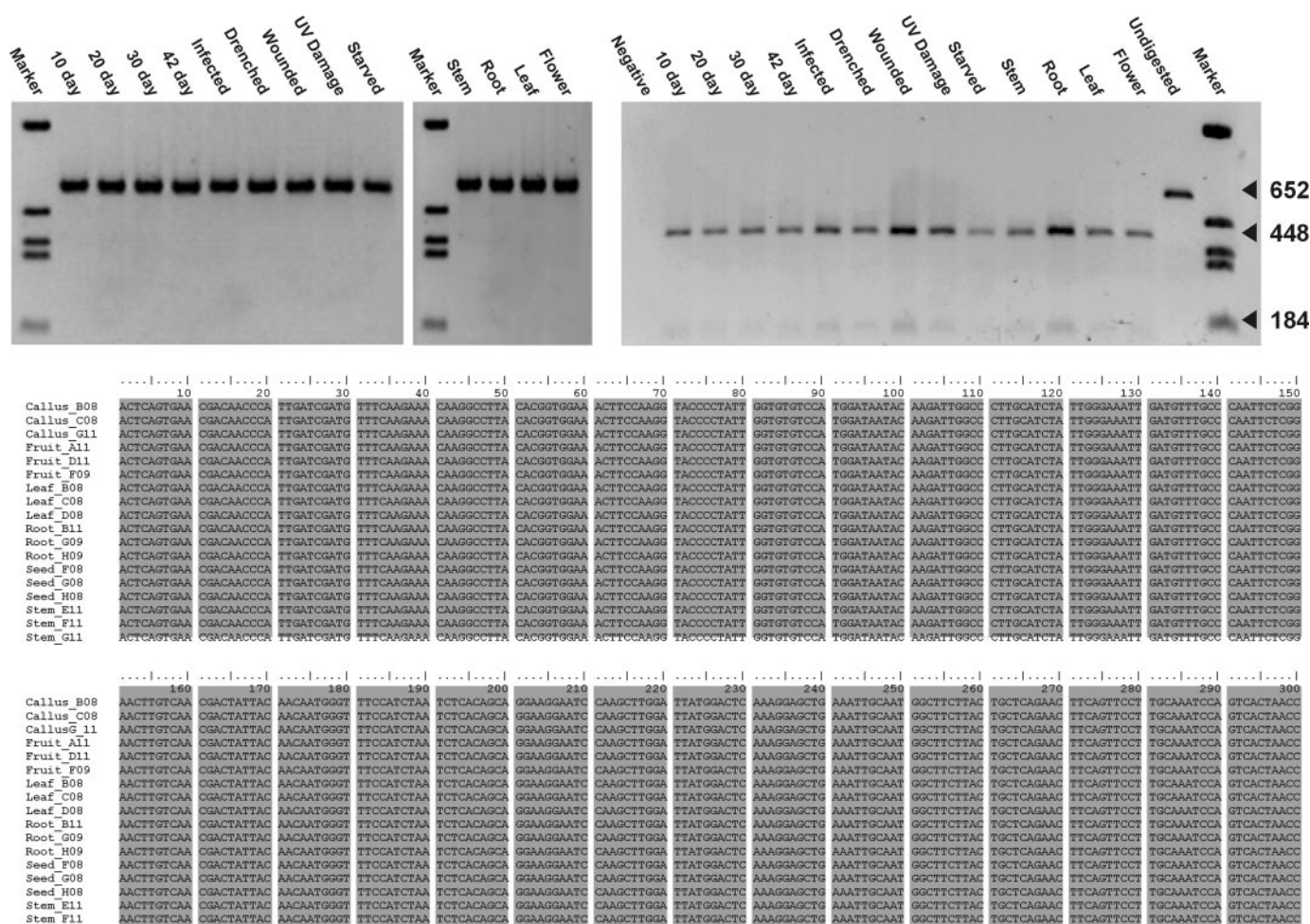


FIGURE 7. Survey of major tomato PAL mRNA sequences of different organs or plant growth under different conditions. Cellular RNA was extracted from tomato leaves of different ages or alternative organs and plants exposed to various environmental stress as indicated and used to prepare exon 2 cDNA by RT-PCR (left panels) as described for Fig. 3. The PCR-amplified cDNA was digested HinfI endonuclease for restriction fragment length polymorphism analyses (right panel) or cloned in pTZ19R for sequence analysis. Representative portions of the cloned sequences from different tomato organs are aligned below to indicate the complete identity.

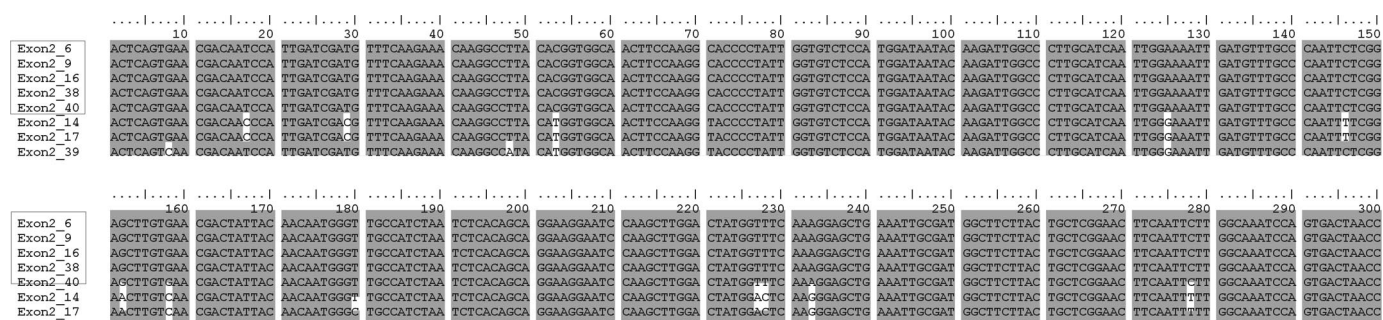


FIGURE 8. Comparison of potato leaf PAL mRNA sequences. Exon 2 cDNAs were prepared from potato mRNA as described in Fig. 3, cloned in pTZ19R and sequences of randomly picked clones were determined by automated sequence analysis. Representative portions of the sequences are compared and grouped for maximum identity; a major group of identical sequences is indicated by the closed box on the left. Shared nucleotides are indicated by shading.

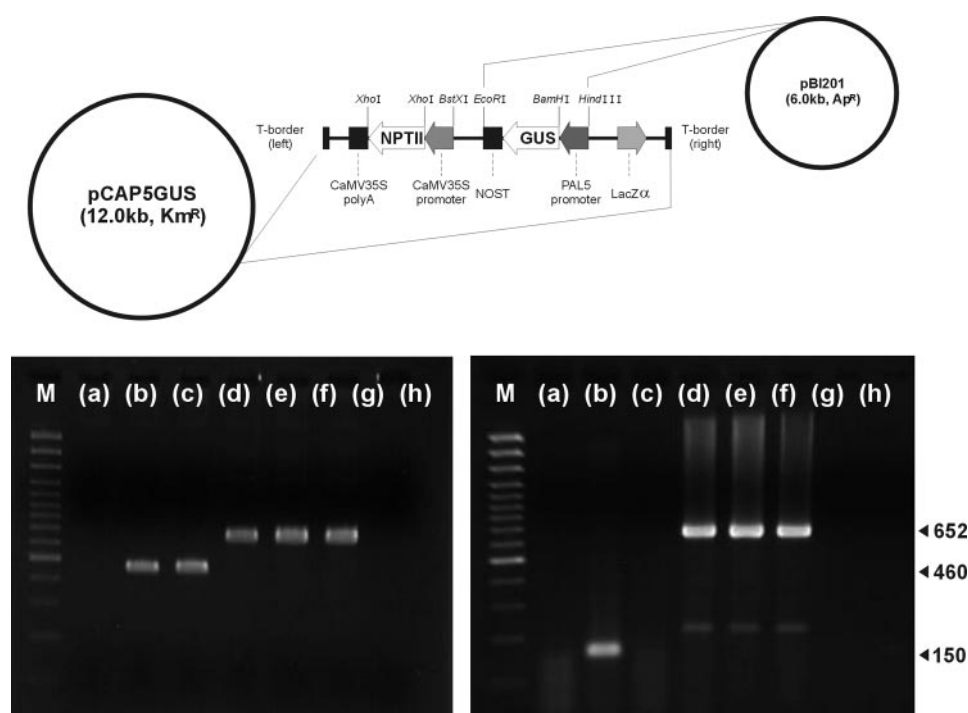
Expression of the GUS reporter sequence under the control of the tomato PAL5 promoter was evaluated by RT-PCR amplification using a GUS gene-specific primer (5'-GTAACATA-AGGGACTG-3') beginning seven bases downstream of the translation start site for reverse transcription and the PAL transcript-specific primer (5'-CACTAATCATAGTTTCACAAC-3') beginning 143 bases upstream of this site. Only a short GUS sequence was included to avoid false negatives caused by any

unanticipated GUS RNA instability. Again, the 150-bp PCR-amplified fragment was detected readily by gel electrophoresis when the transcript was present (Fig. 9, lane b, right gel). At the same time the endogenous PAL gene expression also was assessed as a positive control using exon two specific primers. As summarized in the tabulated results included in Fig. 9, based on these assays in two separate trials beginning with more than 100 cotyledons, about 10% of them (15 and 13, respectively)

whereas neither the transgene in the transcribing plant or the active host plant PAL promoter (PAL_E) were subject to cytosine methylation.

DISCUSSION

Although most plants contain a small number of PAL gene loci, tomato contains a surprisingly large family. Hybridization analyses suggest ~26 copies/diploid genome; at least 18 different sequences were identified in one tomato cultivar in the course of this study. Despite this large gene family and extensive duplication of some of its members, the BAC library analyses clearly show that the genes are not clustered extensively in tandem as might be anticipated with such extensive duplication. Instead they appear widely dispersed in the tomato genome. More surprising was the number of these genes that are expressed. Again, in other plants, many of the PAL gene loci have been shown to be expressed, with different family members often responding to alternative environmental stimuli. Even in this study, which included a preliminary examination of the relatively large PAL gene family in potato, the expression of several different members could be detected readily. In this regard the expression of only a single PAL gene sequence in tomato stands out as a very unusual observation. Based on the extensive sequence comparisons in this study, most of the gene sequences appear



	Cotyledons	Transformants	Expressed mRNA
Trial 1	300	15	1
Trial 2	120	13	1

FIGURE 9. Expression of PAL5-GUS fusion transgenes in regenerated transformed tomato plants. A plant transformation binary vector, pCambia2300, containing a tomato PAL5 gene promoter (8) fused to a GUS reporter sequence and a NOS gene polyadenylation site, which was subcloned from a previously reported construct, pBI201 (29), as indicated in the upper schematic (pCAP5GUS), was used to transform *Agrobacterium*, LBA44044/pAL4404. Tomato cotyledons were then transformed by co-cultivation and plants were regenerated as described under "Experimental Procedures." Transformation was confirmed by PCR amplification (left panel) using PAL gene/GUS gene-specific primers (lanes a–c) or PAL5 gene-specific primers as positive controls (lanes d–f) and reactions without DNA template as negative controls (lanes g and h, respectively). Transgene expression was assayed by RT-PCR (right panel) using PAL gene/GUS gene-specific primers (lanes a–c). PAL5 gene-specific primers also were used as positive controls (lanes d–f), and reactions without reverse transcriptase were included (lanes g and h, respectively) as negative controls. The size of the amplified fragments, as determined with markers (lanes M), are indicated on the right. The results of two independent trials are tabulated below.

were transformed successfully and regenerated as transgenic plants. Of these only one in each trial was found to express the chimeric sequence. Clearly the gene silencing that is so evident in normal plants remained strikingly active with respect to newly introduced PAL gene promoters.

Because all of the PAL genes were very similar in sequence, post-transcriptional gene silencing appeared an unlikely mechanism to explain the expression of a single family sequence. Because upstream features can be more variable and positional effects are possible, the transformed plants were examined further for differences in DNA methylation using bisulfite conversion genomic sequencing (19). As shown in Fig. 10, when plants with silent transgenes (pCAPSG_S1 and pCAPSG_S2) were compared with the plant expressing the GUS RNA sequence under an active PAL promoter pCAPSG_E, a difference in DNA methylation clearly was evident. Both the silent transgenes revealed a mixed pattern of methylation at three sites,

to encode functional PAL protein of normal length, but even if a number are considered to be pseudogenes, the degree of underutilization remains striking. Other studies on the phenylalanine ammonia lyase enzyme in tomato, although limited, are consistent with the current analyses. Fungal inoculations of tomato cell cultures have been reported not to result in multiple forms of the enzyme (32).

Such an observation raises at least two immediate and closely related questions about the nature of PAL gene expression in tomato. Why does differential PAL gene expression not occur in this plant species, and what mechanism restricts gene utilization to such a dramatic degree? In the first instance at least, one explanation is provided by a previous study of PAL 5 gene expression in tomato (9). When sites of RNA initiation were examined using a nuclease protection assay, two distinct transcript initiation sites were identified that appeared associated with alternate TATA sequence ele-

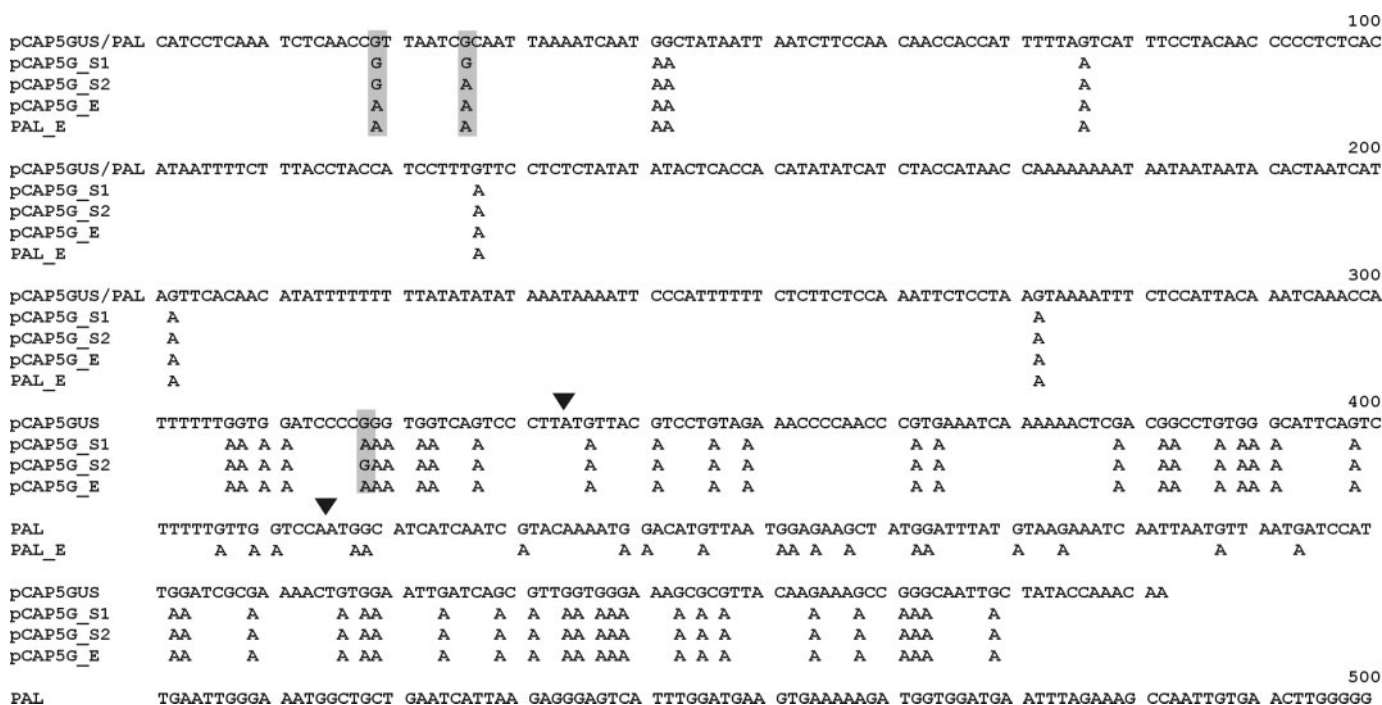


FIGURE 10. A comparison of methylated cytosine residues in the upstream region of transgenes or an expressed PAL gene sequence in tomato plants transformed with pCAP5GUS. Tomato plants were transformed with pCAP5GUS and regenerated as described in Fig. 9; nuclear DNA was prepared from leaves, digested with EcoRI endonuclease, and treated with sodium bisulfite as described under "Experimental Procedures." The upstream regions were PCR-amplified using PAL gene or transgene-specific primers, cloned in the pTZ19R vector, and subjected to DNA sequence analyses. The bisulfite-treated coding strand sequences for two unexpressed transgenes (pCAP5GUS_S1 and pCAP5GUS_S2), an expressed transgene (pCAP5GUS_E), and the expressed host plant PAL gene (PAL_E) are compared together with untreated pCAP5GUS and PAL gene DNA. The common untreated sequence is identified as pCAP5GUS/PAL, whereas the differing regions are identified as pCAP5GUS and PAL, respectively. Protein initiation sites are indicated by the arrowheads; methylated positions in the complementary strand are indicated by shading.

ments. These sites responded differently to environmental stimuli, one acting largely as a constitutive promoter and the other varying significantly with environmental change, much as has been reported with alternate gene members in other plants. Given that, as shown in the present study, the tomato PAL genes are highly underutilized, it appears that in this instance alternative regulatory pathways have been focused on a single gene rather than alternate members of the gene family. Differential regulation of gene expression has not ceased but instead is directed at a single transcript.

Although the current study does not identify a specific reason for the PAL gene underutilization, the transformation analyses do demonstrate a very active silencing mechanism that appears to rule out simple genetic drift as a cause. As indicated in Fig. 9, when new active PAL gene promoters are introduced, they are very efficiently silenced, much as is the large number of normal family members. Of 28 transformants selected in two separate experiments, only two were found to be actively transcribed with more than 90% of the new genes already silenced upon plant regeneration. The actual mechanism is unclear. Although RNA interference has been implicated in many instances of gene regulation or post-transcriptional gene silencing (for reviews see Refs. 33 and 34), this is unlikely in the present case because the sequences seem too similar to allow only the PAL5 sequence to escape a parallel degradation. Furthermore, all methods (restriction fragment length polymorphism, RT-PCR, and cDNA sequencing) failed to detect additional transcripts, an observation that argues strongly against control by any form of

post-transcriptional degradation. DNA methylation (for reviews see Refs. 35–38) or some other nuclear phenomenon appeared more likely. As shown in Fig. 10, analyses of DNA cytosine methylation based on bisulfite conversion genomic sequencing strongly support DNA methylation as the probable cause of gene silencing. Whereas the method remains unclear and will require more detailed comparisons of promoter regions, possibly also including chromosomal positioning, the results identify gene methylation as a significant differentiating factor.

Returning to the initial impetus for this study, namely, the evolution of the PAL gene family in tomato, the results identify an unanticipated feature, an unusually large expansion of the family members, many of which appear to be highly and efficiently silenced. Is unusually active duplication of genes counterbalanced by a more aggressive silencing mechanism? Furthermore, duplication would normally be expected to result in tandemly arranged gene structures, but the PAL genes are, instead, widely dispersed in the tomato genome. These intriguing questions remain unanswered and invite detailed investigation. In the interim, the present study underlines the very large size of this important gene family in tomato and the unusual way in which regulation is focused on a single member to control phenylalanine lyase expression for normal use and in response to environmental stimuli.

Acknowledgments—We are grateful to K. Haq and J.-J. Yu for excellent technical assistance.

REFERENCES

- Ku, H.-M., Vision, T., Liu, J., and Tanksley, S. D. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9121–9126
- Lynch, M., and Conery, J. S. (2000) *Science* **290**, 1151–1155
- Moore, R. C., and Purugganan (2005) *Curr. Opin. Plant Biol.* **8**, 122–128
- Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G., and Tanksley, S. D. (2002) *Plant Cell* **14**, 1441–1456
- Cramer, C. L., Edwards, K., Dron, M., Liang, X., Dildine, S. L., Bolwell, G. P., Dixon, R. A., Lamb, C. J., and Schuch, W. (1989) *Plant Mol. Biol.* **12**, 367–383
- Kumar, A., and Ellis, B. E. (2001) *Plant Physiol.* **127**, 230–239
- Joos, H. J., and Hahlbrock, K. (1992) *Eur. J. Biochem.* **204**, 621–629
- Lee, S.-W., Robb, E. J., and Nazar, R. N. (1992b) *J. Biol. Chem.* **267**, 11824–11830
- Lee, S.-W., Heinz, R., Robb, J., and Nazar, R. N. (1994) *Eur. J. Biochem.* **226**, 109–114
- Hoaglund, D. R., and Arnon, D. (1950) in *Circular of the California Agricultural Experiment Station, Number 347*
- Cushman, J. C. (1995) *Methods Cell Biol.* **50**, 113–128
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Heinz, R., Lee, S. W., Saparno, A., Nazar, R. N., and Robb, J. (1998) *Physiol. Mol. Plant Pathol.* **52**, 385–396
- Marchuk, D., Drumm, M., Saulino, A., and Collins, S. F. (1991) *Nucleic Acids Res.* **19**, 1154
- Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
- Gerad, G. F. (1987) *Focus* **9**, 5–6
- Fillatti, J. J., Kiser, J., Rose, R., and Comai, L. (1987) *BioTechnology* **5**, 726–730
- Stachel, S. E., Messens, E., Montagu, M. V., and Zambryski, P. (1985) *Nature* **318**, 624–629
- Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 1827–1831
- Lee, S.-W., Nazar, R. N., Powell, D. A., and Robb, J. (1992a) *Plant Mol. Biol.* **18**, 345–352
- Budiman, M. A., Mao, L., Wood, T. C., and Wing, R. A. (2000) *Genome Res.* **10**, 129–136
- Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A., and Mathur, E. J. (1991) *Genetics* **180**, 1–6
- Liang, X., Dron, M., Schmid, J., Dixon, R. A., and Lamb, C. J. (1989a) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 9284–9288
- Liang, X., Dron, M., Cramer, C. L., Dixon, R. A., and Lamb, C. J. (1989b) *J. Biol. Chem.* **264**, 14486–14492
- Liao, Y. C., Li, H. P., Kreuzaler, F. M., and Fischer, R. (1996) in *Plant Gene Register* PGR96–PGR102
- Lois, R., and Hahlbrock, K. (1989a) *Biosci. Z. Naturforsch* **47**, 90–94
- Lois, R., Dietrich, A., Hahlbrock, K., and Schultz, W. (1989b) *EMBO J.* **8**, 1641–1648
- Wanner, L. A., Li, G., Ware, D., Somssich, I. E., and Davis, K. R. (1995) *Plant Mol. Biol.* **27**, 327–338
- Thompson, G. A., Boston, R. S., Liznik, L. A., Hodges, T. K., and Larkins, B. A. (1990) *Plant Mol. Biol.* **15**, 755–764
- Holsters, M., D., d. W., Depicker, A., Messens, E., Van Montagu, M., and Schell, J. (1978) *Mol. Gen. Genet.* **163**, 181–187
- Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J., and ASchilperoort, R. A. (1983) *Nature* **303**, 179–180
- Bernards, M. A., and Ellis, B. E. (1991) *Plant Physiol.* **97**, 1494–1500
- Morris, K. V. (2005) *Cell. Mol. Life Sci.* **62**, 3057–3066
- Yu, H., and Kumar, P. P. (2004) *Plant Cell Reports* **22**, 167–174
- Bender, J. (2004) *Annu. Rev. Plant Biol.* **55**, 41–68
- Finnegan, E. J., Genger, R. K., Peacock, W. J., and Dennis, E. S. (1998) *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 223–247
- Jacobsen, S. E. (1999) *Curr. Biol.* **9**, R617–R619
- Kalisz, S., and Perugganan, M. D. (2004) *Trends Ecol. Evol.* **19**, 309–314